BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 2: e25176
doi: 10.3897/biss.2.25176

OPEN ACCESS

Conference Abstract

Data Quality in Data Exchanges: a Tri-Part Approach in the French Information System on Nature and Landscapes

Remy Jomier‡, Solène Robert‡

‡ Natural History Museum, Paris, France

Corresponding author: Remy Jomier (remy.jomier@mnhn.fr), Solène Robert (solene.robert@mnhn.fr)

Received: 21 Mar 2018 | Published: 18 May 2018

Citation: Jomier R, Robert S (2018) *Data quality in data exchanges: a tri-part approach in the French Information System on Nature and Landscapes*. Biodiversity Information Science and Standards 2: e25176.
https://doi.org/10.3897/biss.2.25176

Abstract

As part of the Biodiversity Information System on Nature and Landscapes (SINP), the French National Natural History Museum has been appointed to develop biodiversity data exchanges by the French ministry in charge of ecology. Given there are, quite literally, thousands of different sources, such a development brings into question the underlying quality of data. To add complexity, there can be several layers of quality: one being appraised by the producer himself, one by a regional node, and one by the national node.

The approach to quality issues was addressed by a dedicated working group, representative of biodiversity stakeholders in France. The resulting documents focus on core methodology elements that characterize a data quality process for taxon occurrences only in the first instance (It may be extended to habitats, geology, etc. in the near future).

Three processes are covered, how to ensure:

- data conformity by checking for the presence of compulsory elements or that a given attribute is of the right type,
- data consistency by checking information versus other information (for example, an end date has to be later than a start date),

- and scientific validation, through either manual (use of expertise) or automated (comparison with knowledge databases) means, or even a combined approach that provides users with a quality appraisal of said data.

Within the SINP, only data that has passed conformity and consistency tests can be exchanged with any and all types of validation levels. For example, should there be no expert existing on a specific taxon group, unvalidated data can be shared.

For scientific validation, two processes are used, one automatic that uses several criteria such as comparison with a national taxonomic reference database (TAXREF), and with species reference maps. The combination of all these elements can be used to automatically flag data for a second, deeper, manual process that allows for further scrutiny in order to reach a conclusive evaluation. This allows experts to work only on “doubtful” data, thus saving time.

In the future, other criteria that are currenlty used with the manual approach, such as for example congruity, data scarcity on a given species, determination difficulty, existence of associated proof (specimen, picture…), knowledge of the ability of the observer, databases on most frequent determination errors etc., could be added to the automatic process.

Some elements must be included in the data to allow for comprehensive testing, and have been included in a national data standard so that the result of the validation process can be shared with users, allowing them to judge how the data is fit for their use.

The presentation will deal with how such a work was undertaken and how conformity, consistency and scientific validation have been treated and issues solved by the workgroup. For example, there could be a 40 million data record backlog. The presentation will also show how the required elements could be integrated into the French national standard.

Keywords

data quality, data exchange, scientific validation, biodiversity

Presenting author

Rémy Jomier